

## 6 Conclusões

Neste trabalho apresentou-se como se ter confiança em um agente cognitivo com crenças, desejos e intenções explicitamente, tornando tal componente mais um estado mental para um agente inteligente.

Primeiramente, enfatizou-se o porquê de se ter um modelo de confiança em um agente que interage com outros em um sistema multi-agentes aberto e por que a confiança difere da crença.

Para se colocar confiança no modelo, inicialmente se acrescentou a confiança à definição de um agente BDI. Posteriormente, colocou-se como a confiança interage com os outros componentes do agente durante o funcionamento de seu *loop* de controle em suas várias versões.

Na formalização do sistema em lógica modal, em primeiro lugar foi formalizada a entrada e a saída de agentes do sistema e diversas propriedades de sistemas multi-agentes dependendo de como ocorre a entrada, a saída e a permanência de agentes no sistema.

Em seguida, foi formalizado o modelo de agente com confiança em lógica modal e usando a semântica de mundos possíveis. Para isso, estendeu-se a linguagem LORA (Woo00). Nessa extensão, para as trocas de mensagens, as delegações de tarefas, e a entrada e a saída de agentes do sistema, foram definidas ações específicas, que são as ações realizadas pelos agentes no ambiente. Também, permitiu-se a criação de novas ações específicas, algo importante para se descrever propriedades das interações entre agentes.

Outras adições à lógica foram predicados específicos para falar dos diferentes tipos de confiança que um agente pode ter em seus parceiros. Também foi incluída uma nova modalidade que se refere às mensagens enviadas por um agente. Tem-se ainda a adição do conceito de capacidades para a realização de uma tarefa ou dizer algo. Com a modificação da sintaxe da lógica e o fato de o sistema ser aberto, ao contrário do que acontece em (Woo00), a sua respectiva semântica teve que ser modificada para englobar as extensões feitas.

Mais ainda, foram definidas propriedades relacionando a atualização da confiança com a atualização das crenças de um agente, quando e como se deve

atualizar o grau de confiança, e quando se deve interagir com outro agente.

Em seguida, foram formalizados os predicados *concis()*, que diz quando um plano é consistente com as crenças e as intenções, e *possivel()* e *impossivel()*, que definem quando um conjunto de intenções é possível ou não, dadas as crenças do agente.

Finalizando, foram apresentados dois exemplos práticos simples para a lógica quanto à avaliação de vendedores da Amazon.com Marketplace e para a escolha de um vendedor dadas várias opções. Esses exemplos mostram, de maneira simples, como a lógica descrita neste texto pode ser útil para a definição formal e validação de propriedades, políticas e recomendações para um MAS aberto com confiança.

## 6.1

### Decisões de Projeto

Para a linguagem usada para os agentes, optou-se por apenas estender LORA para se ter confiança (predicados de confiança), entrada e saída de agentes como ações específicas, ações para o envio de mensagens e a delegação de tarefas, a possibilidade de se formular novas ações específicas, uma modalidade para dizer que um agente disse algo e a noção de capacidades para a realização de uma ação ou para a emissão de uma mensagem.

No caso da entrada e da saída de agentes, não é especificado o instante de tempo em que ele entra no sistema. O mesmo ocorre com as mensagens enviadas entre os agentes, as delegações e as demais ações. Isso acontece porque em LORA a noção de tempo não é explícita na linguagem lógica. Embora ela tenha um componente temporal, ela não é uma linguagem para lidar com a noção de tempo real. Se for necessário saber o instante exato no qual ocorrem as ações, por exemplo, seria interessante acrescentar a lógica essa noção de tempo. Todavia, isso tornaria muito mais difícil a implementação de um validador de modelos usando essa lógica. Afinal, com tempo real tem-se uma explosão do número de estados (Cla99).

No presente texto não se modelou a possibilidade de um agente mentir para outro, e como o outro poderia detectar isso e o que faria em resposta. Acredita-se que para isso ser modelado seria necessário um predicado que indicasse quando o agente está mentindo.

Quanto ao fato de as mensagens enviadas não mostrarem o destinatário, isso foi feito para se usar uma sintaxe semelhante a BAN. Como dito na seção 5.3, essa informação pode ser passada em uma camada mais acima do protocolo.

Por fim, considerou-se que a atualização das crenças e do modelo de confiança são instantâneas, o que na prática pode não ocorrer.

## 6.2 Trabalhos Futuros

Primeiramente, nas fórmulas das modalidades de crenças, desejos e intenções, e nos predicados de confiança, seria interessante, ao invés de se ter apenas agentes individuais, tais como  $i$  e  $j$ , ter conjuntos de agentes. Ou seja, as modalidades e os predicados citados poderiam ser estabelecidos para grupos de agentes e não apenas para agentes individuais. Isso refletiria o estado mental comum de grupos de agentes. Em LORA isso é permitido para crenças, desejos ou intenções mútuas (Woo00). Seria interessante estender o conceito para a confiança, embora ela não seja uma modalidade. Com isso, seria possível fazer inclusive inferências sobre o que há nos estados mentais de um grupo baseado nos dos agentes individuais e vice-versa.

Outra modificação na abordagem é que, da mesma forma que se criou um estado mental para confiança, talvez fosse interessante criar um estado mental sobre o que os agentes dizem para os outros, já que tal tipo de informação é encontrada em muita quantidade nas crenças do agente. Afinal, os agentes, ao interagir, trocam muitas mensagens. Nesse caso o estado mental seria mesmo uma modalidade, tal como foi representada nesse texto. Esse estado mental teria todas as trocas de informação que o agente consegue capturar do ambiente e que lhe são relevantes. Afinal, tanto no caso das crenças quanto das mensagens trocadas, ele não precisa armazenar informações que não lhe interessam.

Outro estado mental que poderia ser adicionado ao agente é o conhecimento (Woo00). Conhecimento é diferente de crença porque uma vez que um agente sabe algo, ele não deixa de saber. Logo, além do  $KD45$  que vale nas crenças, o axioma  $T$  vale no conhecimento, fazendo dele uma modalidade  $S5$ . Seria interessante estabelecer relações entre as crenças e o conhecimento, e os demais estados mentais do agente, incluindo a confiança.

Seria interessante também a modelagem de fatos, que são verdades, os quais poderiam ser sobre propriedades do ambiente, ações e comportamento de agentes, por exemplo. Isso permitiria modelar precisamente quando um agente está mentindo, falando a verdade, cometendo omissões em suas mensagens, etc.

Mais ainda, assim como a confiança é algo probabilístico no caso não-binário, poderia se fazer o mesmo com as crenças. Ou seja, um agente, ao invés de simplesmente colocar uma representação do que ele percebe do ambiente em

suas crenças, ele poderia atribuir-lhe um valor probabilístico, que significaria a chance de ele estar tendo uma percepção correta do que foi analisado. Afinal, nem sempre um agente, assim como os seres humanos, conseguem perceber o ambiente corretamente e, em alguns casos, poderiam ter consciência de suas limitações. Resumindo, as crenças poderiam ter graus variados de “achismos”, que é o que o agente “acha” que está percebendo do seu ambiente. Note-se que isso não é o mesmo do que confiar em algo do ambiente, já que supõe-se que apenas a percepção do agente em algo do seu ambiente esteja em questão. Ou seja, o agente não está apostando que o ambiente vai se comportar de uma determinada forma e sim está fazendo suposições com maior ou menor grau de certeza sobre o estado atual do ambiente.

As intenções também poderiam ter um grau associado, mas, neste caso, seria um valor que indicaria o quanto o agente está comprometido a ela. Quanto maior o grau, mais importante é o alcance da intenção pelo agente. Isso é semelhante ao que ocorre na vida real com as pessoas, onde determinados objetivos a serem alcançados são mais importantes do que outros. De posse desses valores, um agente poderia escolher o quanto ele pode se comprometer com cada intenção de acordo com os seus recursos disponíveis.

Também pretende-se implementar uma ferramenta CAV (*Computer Aided Validation*), que poderia ser um *model-checker* onde se entraria um modelo de sistema, com os agentes que podem entrar e sair e como eles interagem entre si, e fórmulas que dizem a respeito de propriedades internas do agente (incluindo seus estados mentais), do sistema e das interações e ações dos agentes. Essas propriedades poderiam ser algumas das fórmulas citadas na seção 5.7. Com isso, o validador analisaria as fórmulas de encontro ao modelo verificando se o mesmo satisfaz as propriedades desejadas. Dessa forma, pode-se verificar se determinado agente que se comporta adequadamente vai ser confiável em alguma situação, por exemplo. Ou então, verificar se um agente vai deixar de ser confiável em algum estado futuro após agir maliciosamente.

Adicionalmente, deseja-se investigar mais propriedades da lógica exposta, como a sua completude, grau de complexidade e decidibilidade. Ainda com relação a lógica, pretende-se explorar a possibilidade de se ter um sistema multi-agentes com a lógica modal  $GL$  apresentada na seção 5.2, onde  $\Box_i\varphi$  seria lido como o agente  $i$  é capaz de demonstrar (provar)  $\varphi$ . Isso poderia vir a ser um novo estado mental do agente, que incluiria o que ele pode provar. Afinal, mesmo ele tendo um estado mental de conhecimento, ele não é capaz de demonstrar tudo o que sabe. Já o contrário, saber tudo o que é capaz de demonstrar, tem que ser verdadeiro para que o agente seja racional.

Ainda, deseja-se fazer uma comparação detalhada sobre a escolha de

planos para se atingir uma intenção com o operador *“bring-it-about”* (Ama00, Seg89) e tentar fazer um mapeamento de uma abordagem para a outra. Esse operador poderia ser útil para a inclusão de tempo real nas ações executadas pelos agentes na lógica do presente texto, tal como acontece em (Ama00).

Outro tabalho futuro interessante é a comparação da abordagem aqui apresentada para a modelagem de confiança com os modelos probabilísticos para confiança, tal como os que usam cadeias de Markov (Huy05, Mul05, Smi05).

Por fim, deseja-se explorar um contexto prático onde o modelo de confiança descrito neste texto possa ser implementado de maneira satisfatória e eficiente.